

Extremist Cultures Online

INTS1302 NAVIGATING DIGITAL WORLDS, WEEK 6

Assessment 1 due Friday

THE TUTORIAL IS YOUR WORKING SESSION

- Assessment 1 is due Friday 4 September.
- This week's tutorial is a working session for it.

Assessment 1 is due Friday 4 September.
This week’s tutorial is a working session for it.

And now
for something
completely different...



John Cleese, Monty Python's Flying Circus (BBC).

Content Warning

CRITICAL ANALYSIS, NOT ENDORSEMENT

Content Warning. *The content of this report necessarily engages with themes that are extreme and anti-social. Examples of extreme, hateful, and violent language are included in order to faithfully reproduce the data collected. Some racial and religious slurs as well as an obscenity have been redacted.*

([Ballsun-Stanton et al., 2020](#))

- the warning is the report’s own, covering the assigned reading
- the lecture shows the report’s descriptions and patterns, not its raw examples

Support is available

SERVICES YOU CAN CONTACT DURING OR AFTER THIS LECTURE

- You can step out at any time, or ask me to pause or skip material.
- 1800 CARE MQ (1800 2273 67) connects you with support anytime, any day.
- Lifeline is available any time on 13 11 14.

students.mq.edu.au → **Support** → **Wellbeing** Counselling and Wellbeing, and the confidential CARE Form self-referral.

You can step out at any time, or ask me to pause or skip material.

1800 CARE MQ (1800 2273 67) connects you with support anytime, any day.

- counselling, medical, accessibility, welfare and advocacy services in one place
- a quick, confidential self-referral through the CARE Form

Lifeline is available any time on 13 11 14.

One investigation at every scale

FROM SINGLE POSTS TO POPULATION CLAIMS, AND WHAT EACH STEP CAN SHOW

- In 2020 we mapped online right-wing extremism in New South Wales for the state government.
- One government question became evidence, findings, recommendations, and what government did next.
- Evidence lives at many scales, from a single post to whole platforms.
- A claim's scale must fit the scale and construction of its evidence.
- A corpus is the whole collection of posts gathered for one analysis, and corpora is its plural.

posts → conversations → accounts → networks → corpora → platform landscapes → the population (beyond this evidence)

In 2020 we mapped online right-wing extremism in New South Wales for the state government.

- commissioned by the state, published openly on Zenodo
 - DOI 10.5281/zenodo.4071472, version 1.0.1, CC BY-ND ([Ballsun-Stanton et al., 2020](#))

- a real team, real deadlines, a real government reader

One government question became evidence, findings, recommendations, and what government did next.

- chronology as the lecture’s spine

- one stage per section, running question, evidence, findings, recommendations and uptake

Evidence lives at many scales, from a single post to whole platforms.

- a single post readable in seconds, a million only as patterns

- each rung of the map a different kind of evidence

- up the map more coverage and less context, down the map the reverse

A claim’s scale must fit the scale and construction of its evidence.

- the week’s governing rule

- big claims need evidence built to carry them

- modest evidence still earns modest claims

A corpus is the whole collection of posts gathered for one analysis, and corpora is its plural.

- the whole collection gathered for one analysis

- two corpora anchor this study

Displayed scale map

- posts, conversations, accounts, networks, corpora, platform landscapes, and the population claims beyond the evidence

- previewed here, and returning complete at the recap

March 2019, Christchurch

AN ATTACK MEDIATED ONLINE, BY AN ATTACKER FROM NEW SOUTH WALES

- In March 2019 a terrorist murdered 51 people at mosques in Christchurch, New Zealand.
- The attack was streamed and shared online, and the attacker came from New South Wales.
- The New South Wales Government needed to understand its own online extremist environment.

In March 2019 a terrorist murdered 51 people at mosques in Christchurch, New Zealand.

- 15 March 2019, two mosques, 51 dead

- no imagery of the attack in this lecture

The attack was streamed and shared online, and the attacker came from New South Wales.

- livestreamed and recirculated across platforms

- New Zealand’s Royal Commission traced the online life of the attack ([Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on 15 March 2019, 2020](#))

- “This breadth of scope was considered pertinent in the wake of the March 2019 mass casualty terrorist attack in Christchurch, New Zealand, by an attacker originating from NSW.” ([Ballsun-Stanton et al., 2020, p. 1](#))

The New South Wales Government needed to understand its own online extremist environment.

- an environment little studied before 2019

- “wide exploratory research into the online RWE movement in NSW, a milieu that has been little studied” ([Ballsun-Stanton et al., 2020, p. 1](#))

- the question was local, what is here in this state

The commission

A MAPPING STUDY BY A TEAM INCLUDING YOUR LECTURER

- The Department of Communities and Justice commissioned a mapping study in 2019.
- We delivered it, academics from Macquarie University and Victoria University.
- We studied communities we opposed, which raises the standard for careful description.

The Department of Communities and Justice commissioned a mapping study in 2019.

- Department of Communities and Justice, Countering Violent Extremism program
 - “Funding was provided by the Department of Communities and Justice (NSW), Countering Violent Extremism (CVE) Program.” ([Ballsun-Stanton et al., 2020, p. 1](#))
- July 2019 to February 2020, collection August to November 2019
 - “The project was conducted between July 2019 and February 2020 with data collection occurring from August to November 2019.” ([Ballsun-Stanton et al., 2020, p. 1](#))

We delivered it, academics from Macquarie University and Victoria University.

- academics from Macquarie and Victoria University
- delivered under contract with ethics approval

We studied communities we opposed, which raises the standard for careful description.

- opposition to the ideology stated openly in the report
- fairness as method, not sympathy

Three research questions

WHAT THE METHODS MUST BE ABLE TO ANSWER

- Three broad questions: the environment's nature, its narratives, its risk.
- The methods that follow must be able to answer them.

1. What is the nature of the online RWE environment in NSW?
2. How are themes and narratives framed in different online contexts in order to mobilise support?
3. What level of risk does the online RWE environment pose?

([Ballsun-Stanton et al., 2020, p. 1](#))

Three broad questions: the environment’s nature, its narratives, its risk.

- nature, narratives, risk as three angles on one environment
- broad on purpose, exploration over hypothesis testing
 - “The research areas were framed as broad questions to facilitate wide exploratory research” ([Ballsun-Stanton et al., 2020, p. 1](#))

The methods that follow must be able to answer them.

- every later method judged against these questions
- a method that cannot answer them wastes the client’s money

Displayed research questions

- the three questions on screen exactly as printed at page 1
- the map’s first descent begins from these questions

The Gab corpus

FROM SEED ACCOUNTS TO A CONNECTED MILIEU

- Gab is a small social platform with minimal moderation, popular with the far right.
- I started from supplied seed accounts and followed reciprocal connections outward.
- Snowball sampling finds a connected community, not a population.
- The Gab collection found users interested in Australia rather than Australian users.

Gab is a small social platform with minimal moderation, popular with the far right.

- marketed on free speech and light moderation
 - grew sharply after its July 2019 relaunch drew press coverage
- much smaller than Twitter or Facebook

I started from supplied seed accounts and followed reciprocal connections outward.

- seeds supplied to us, snowball outward from them
- a reciprocity rule, only accounts both followed by and following the targets
- three months of continuous collection

Snowball sampling finds a connected community, not a population.

- finds the connected, misses the unconnected
- indicative, not representative

- “The account snowball targeting strategy is indicative: we seek tightly connected communities of users and extend out, rather than collecting all possible posts from a region and looking for users of interest thereby.” (*Ballsun-Stanton et al., 2020, p. 15*)

The Gab collection found users interested in Australia rather than Australian users.

- interest in Australia, not Australian residence
 - “From our snowball on Gab, we were able to find users interested in Australia, rather than Australian users.” (*Ballsun-Stanton et al., 2020, p. 16*)
- identity and location unverifiable at scale

The Twitter corpus

17,000 POSSIBLE ACCOUNTS, 27 CHOSEN

- On Twitter, 17,000 possible accounts were too expensive to collect.
- We hand-picked 27 accounts and I collected the conversations around them.
- A separate purchase of 300,000 Australian tweets narrowed to 37,442 from users naming a New South Wales location.
- Cost, selection and self-reported location all shape what the sample can show.

On Twitter, 17,000 possible accounts were too expensive to collect.

- 17,000 accounts surfaced by the snowball
- Twitter’s historical data priced beyond the budget
 - “While we initially believed it would be possible to acquire data from the 17,000 users in our snowball, the cost of doing this on Twitter’s One-off Historical PowerTrack Service was prohibitive.” (Ballsun-Stanton et al., 2020, p. 12)

We hand-picked 27 accounts and I collected the conversations around them.

- 27 accounts chosen by hand
- all tweets to and from them, plus retweets and quotes
 - “From this list of 17,000 accounts, we manually selected 27 to get all tweets to and from them, in addition to retweets and quoted tweets.” (Ballsun-Stanton et al., 2020, p. 12)

A separate purchase of 300,000 Australian tweets narrowed to 37,442 from users naming a New South Wales location.

- 300,000 tweets over 366 days from Australian-identified accounts
 - “we purchased 300,000 tweets from a period of 366 days originating from Australian-identified accounts, and narrowed that down to 37,442 tweets from 3,321 users positively identifying from being NSW.” (Ballsun-Stanton et al., 2020, p. 14)

- narrowed by public profile locations naming New South Wales
 - 3,321 users in the final Twitter set

Cost, selection and self-reported location all shape what the sample can show.

- every choice a trade-off among money, time and coverage
- the sample is built, not found

Too big to read

1.36 MILLION POSTS AND THE TURN TO DISTANT READING

- The Gab collection alone holds 1.36 million posts from 23,836 accounts.
- No one can read that many posts one by one.
- Distant reading uses computation to find patterns across an entire corpus.
- Franco Moretti coined the term for studying literature at scale.

The Gab collection alone holds 1.36 million posts from 23,836 accounts.

- collection alone ran three months

- posting outpaces any reader

- “The download of 1,357,391 ‘Gabs’ (posts), from 23,836 accounts interested in and interacting with our seed Australian users took more than three months of continuous operation.” ([Ballsun-Stanton et al., 2020, p. 13](#))

No one can read that many posts one by one.

- close reading does not scale
 - about two working years at ten seconds a post

- skimming is not analysis

Distant reading uses computation to find patterns across an entire corpus.

- term frequencies, shared links, hashtags, timing

- the computer counts, the researcher interprets

Franco Moretti coined the term for studying literature at scale.

- Moretti’s provocation from literary studies

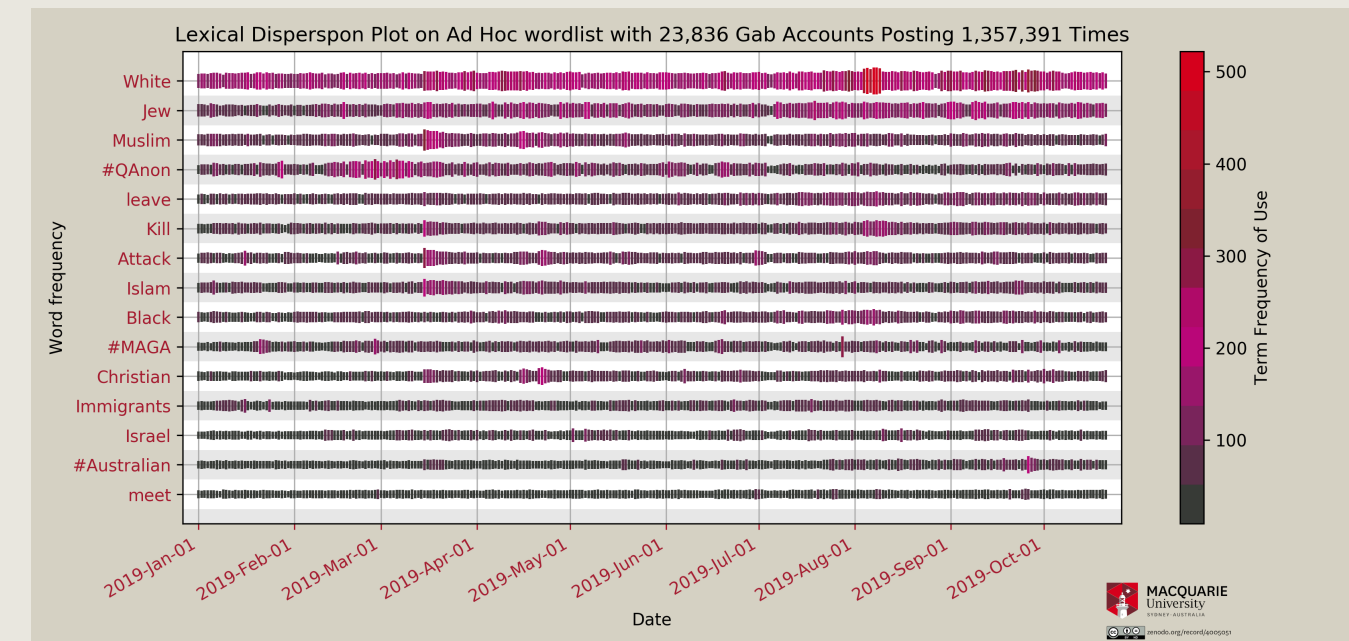
- “we know how to read texts, now let’s learn how not to read them. Distant reading: where distance, let me repeat it, is a condition of knowledge” ([Moretti, 2000, p. 57](#))

- understanding more of literature by reading fewer books closely

Reading the pattern

THE PLOT SHOWS WHEN TERMS MOVE, NOT WHAT THEY MEANT

- Each row is a term. Each mark is a moment it appeared in the corpus.
- A spike shows that a term moved. It cannot say what users meant.
- We went back to the conversations to interpret each pattern.
- Analysis moves both ways between the aggregate and the particular.



The aggregate lexical-dispersion plot, an unaltered analysis output distributed with the report, CC BY-ND ([Ballsun-Stanton et al., 2020](#)).

Each row is a term. Each mark is a moment it appeared in the corpus.

- fifteen terms we chose, an ad hoc wordlist
- group labels, hashtags and plain verbs, nothing needing redaction
 - the report's own output, shown unaltered
 -  "See figures in Appendix 9 for some of the outputs in this process." (*Ballsun-Stanton et al., 2020, p. 15*)
 - the displayed figure is the unaltered standalone Zenodo file, CC BY-ND

A spike shows that a term moved. It cannot say what users meant.

- movement visible, meaning absent
- a spike could be outrage, mockery or news

We went back to the conversations to interpret each pattern.

- back to threads, platform context and human judgement

- the report's own workflow of counting first and then reading

Analysis moves both ways between the aggregate and the particular.

- down for context, up for coverage

- the map from the opening, in action

What each sample can say

TWO CONSTRUCTIONS, TWO DIFFERENT CLAIMS, NO POPULATION ESTIMATE

- The Gab sample shows a connected milieu around chosen seeds.
- The Twitter sample shows selected public conversations from self-identified locations.
- Neither is a random sample, so neither can say how common extremism is in New South Wales.

The Gab sample shows a connected milieu around chosen seeds.

- a milieu connected to chosen seeds

- who talks to whom, not how many think it

The Twitter sample shows selected public conversations from self-identified locations.

- public conversations around 27 chosen accounts

- location self-reported, period fixed

Neither is a random sample, so neither can say how common extremism is in New South Wales.

- no random draw, no population estimate

- two different corpora licensing two different claims

- “we did not perform a purely random sampling of conversations in NSW. This random sampling would be necessary for any statistical analysis of a population.” (*Ballsun-Stanton et al., 2020, p. 14*)

What we found

PLATFORM ENVIRONMENTS, SHARED CONTENT, MODERATION, TWO LEVELS OF RISK

- Platforms sit on a spectrum from low to high risk, set by echo-chamber strength and moderation.
- Different platforms drew on different sources of shared content.
- Two levels of risk: a creeping shift in acceptable discourse, and individuals advocating violence.
- Some findings come straight from the corpora; others rest on published research the report cites.

Platforms sit on a spectrum from low to high risk, set by echo-chamber strength and moderation.

- echo-chamber strength and moderation as the two axes

- “categorised from low through to high against two criteria: the strength of the echo chamber and the level of platform moderation” (Ballsun-Stanton et al., 2020, p. 31)

- Twitter and Facebook low, Gab and the chans high
 - Reddit sits between the poles on both axes

Different platforms drew on different sources of shared content.

- the report lists each dataset’s most-shared external domains
 - the most-frequent external domain lists for the Twitter and Gab datasets (Ballsun-Stanton et al., 2020, p. 34)

- what circulates differs with the platform and its audience

Two levels of risk: a creeping shift in acceptable discourse, and individuals advocating violence.

- creeping threat as the acceptable shifting outward

- “This environment is characterised by narratives that challenge the fundamentals of pluralist liberal democracy through exclusivist appeals to race, ethnicity, nation, and gender.” (Ballsun-Stanton et al., 2020, p. 2)

- violent actors as a minority advocating violence

- “Within RWE communities there are individuals that advocate the use of violence as a tactic for expressing political grievances.” (Ballsun-Stanton et al., 2020, p. 2)

Some findings come straight from the corpora; others rest on published research the report cites.

- corpus claims and literature claims side by side

- the report cites published research for the theory

- Bright on echo chambers, Conway on memes as cover

- “[T]he power of echo chambers may derive from the cognitive tools which they provide people to allow them to hold extreme viewpoints whilst being exposed to dissonant situations” the report quoting Bright and colleagues (Ballsun-Stanton et al., 2020, p. 32)

Sociable and hateful at once

FAIR DESCRIPTION DOES NOT REQUIRE NEUTRALITY

- These communities are genuinely social spaces with shared values and real debate.
- The same spaces are hateful, anti-democratic and risky.
- Fair description reports both truthfully. Neutrality is not required.

These communities are genuinely social spaces with shared values and real debate.

- networks maintained around shared values
- positive experiences for participants
 - “Despite being hateful and extreme, online RWE communities are spaces of sociability for users, where social networks are created and maintained around shared values and norms. They engender positive experiences for those involved in them, but with a revolutionary anti-government agenda hidden behind appeals to nationalism and Australian values.” (Ballsun-Stanton et al., 2020, p. 58)
- real disagreement on the lower-risk platforms
 - “It is important to note that even within echo chambers significant debate and discussions occur.” (Ballsun-Stanton et al., 2020, p. 32)
 - on Gab and the chans, exchanges elaborate rather than oppose
 - “On Gab and the Chans, despite the use of highly offensive language, these exchanges are usually not inherently oppositional as far as the underlying ideology is concerned. Instead they often seek to inform and expand understandings and beliefs” (Ballsun-Stanton et al., 2020, p. 32)

The same spaces are hateful, anti-democratic and risky.

- delegitimising democracy, dehumanising others
 - “at their core these narratives consistently seek to delegitimise liberal democratic government whilst dehumanising others, in particular Muslims, Asians, Jews, women, and liberals” (Ballsun-Stanton et al., 2020, p. 58)
 - both said in the same paragraph of the report
- Fair description reports both truthfully. Neutrality is not required.
- describe accurately, judge openly
 - neutrality would misdescribe the object

Four things the evidence cannot show



MACQUARIE
University
SYDNEY • AUSTRALIA

INTENT, READERSHIP, PREVALENCE, AND CAUSE

- Posts alone cannot separate irony and bragging from real capability and intent.
- Collected accounts cannot show who silently reads.
- Constructed samples cannot estimate how widespread these views are.
- Observed association cannot prove a path to changed belief or violence.

Posts alone cannot separate irony and bragging from real capability and intent.

- bragging, irony and fantasy versus capability and intent
- the report itself calls the separation problematic
 - “Separating examples of bragging, irony, and fantasy from real capability and intent is problematic. This makes identifying potential violent actors from within these communities by their posts alone difficult.” (Ballsun-Stanton et al., 2020, p. 2)

Collected accounts cannot show who silently reads.

- collection saw posters only

- views and lurkers invisible to the APIs
 - ethics bounded collection to what the APIs expose
 - “Ethically, we only interacted with data the platform APIs exposed: we don’t know the rate of views for any given post, nor how many passive users read these posts.” (Ballsun-Stanton et al., 2020, p. 15)

Constructed samples cannot estimate how widespread these views are.

- built samples, no base rate

- no way from milieu size to state prevalence

Observed association cannot prove a path to changed belief or violence.

- association observed, causation unproven

- pathways to violence lie outside the data

The intelligence product

RESEARCH BUILT TO PERSUADE ONE SPECIFIC CLIENT

- The report is an intelligence product written for one client, the New South Wales Government.
- It connects questions, evidence, findings, limits and possible action in one argument.
- Its job is to be usable by a reader who must decide something.

The report is an intelligence product written for one client, the New South Wales Government.

- written to be read by officials, not peers
 - briefed in person to parliament, police and agencies
- plain findings, bounded claims, options for action

It connects questions, evidence, findings, limits and possible action in one argument.

- one argument from question to recommendation
- limits stated as part of the product, not buried

Its job is to be usable by a reader who must decide something.

- usable meaning decision-ready
- credibility spent when claims outrun evidence

What did we recommend?

FIVE AREAS FOR POLICY, AND WHO COULD ACT ON EACH

- Four research priorities and five areas for policy consideration close the report.
 - A useful recommendation names who could act, what could change and who is affected.
 - The report's own areas vary in how explicitly they do this.
1. "Awareness raising for key stakeholders across different levels of government and civil society about the revolutionary and anti-social agenda of right-wing extremism."
 2. "Building awareness about the civic underpinnings of representative liberal democracy and the threat that right-wing extremism poses."
 3. "Expanding current CVE infrastructure provided by the NSW government to individuals and communities vulnerable to right-wing extremism."
 4. "Right-wing extremism exists across urban, regional, and rural NSW, with local government being well positioned to deliver programs in some communities."
 5. "Upskilling front-line workers to recognise the risks associated with right-wing extremism; and providing pathways into CVE intervention programs for individuals identified as being at-risk."

Four research priorities and five areas for policy consideration close the report.

- four research priorities, five policy areas
- awareness, civic understanding, CVE expansion, local government, front-line workers

A useful recommendation names who could act, what could change and who is affected.

- actor, action, affected as the test of a usable recommendation
- vague recommendations vanish inside government

The report’s own areas vary in how explicitly they do this.

- some areas name actors, some name none
 - one area is a statement of position rather than an action
- reading them against the test shows the difference

Displayed policy areas

- the five suggested areas of consideration for policy makers, quoted verbatim ([Ballsun-Stanton et al., 2020, p. 59](#))
- read them with the actor, action and affected test in hand

What happened next

EVIDENCE GIVEN, LANGUAGE CARRIED, NO DIRECT POLICY CHANGE

- The client commissioned further mapping work.
- Team members gave evidence to a New South Wales inquiry and a federal inquiry.
- A parliamentary committee repeatedly cited a submission that drew its description of the problem from our report.
- No chamber mention and no numbered finding directly cited the report.

The client commissioned further mapping work.

- a second mapping phase funded by the same client
 - stated as fact in our federal submission, nothing more claimed
- “The Department of Communities and Justice, NSW have funded our team at Macquarie University to complete a second phase of this mapping research” (PJCIS Submission 12, 2021, para 1.2, captured at docs/inputs/hansard-uptake_2026-08-22/pjcis-submission12-droogan-waldek_2021.pdf)

Team members gave evidence to a New South Wales inquiry and a federal inquiry.

- federal parliamentary hearing, April 2021
 - Waldek and Droogan before the PJCIS on 30 April 2021, Ballsun-Stanton named on the record in absentia
- New South Wales inquiry hearing, February 2026
 - Droogan affirmed and examined before the NSW committee on 18 February 2026
 - the federal inquiry lapsed unfinished at the 2022 election
 - a process can evaporate after the evidence is in

A parliamentary committee repeatedly cited a submission that drew its description of the problem from our report.

- the research network’s submission described the problem in the report’s words
 - “In research reports produced for the NSW Department of Communities and Justice by members of the AVERT Network, RWE in NSW has been described as communities and individuals committed to extreme positions that are pro-white identity” (Submission 27, printed p. 4, footnote 1 citing the 2020 report)
- the committee cited that page repeatedly
 - the one sentence it quoted verbatim traces to a newspaper, not the report
 - the research anchored the definitions, the press supplied the quotable line

No chamber mention and no numbered finding directly cited the report.

- bounded chamber-negative searches by title and surname across both parliaments, recorded in docs/research/week6-hansard-uptake.md

- influence without a citation trail to the floor of parliament
- the strongest honest claim, evidence base rather than policy change

Why write reports nobody acts on?

SHAPING WHAT GOVERNMENT CAN SAY AND SEE

- Influence can be indirect: language, framing and evidence travel without credit.
- The research network carried the report's description into the submission and committee material.
- An intelligence product can matter even when no one acts on it immediately.

Influence can be indirect: language, framing and evidence travel without credit.

- framing travels through submissions into committee language

The research network carried the report’s description into the submission and committee material.

- its description recurs through the committee’s characterisation paragraphs

- in the committee’s characterisation, not its named findings

An intelligence product can matter even when no one acts on it immediately.

- intelligence value lies in readiness

- the state reaches for what already exists

What distance cannot show

ACTIVITY, READERSHIP, AND THE LIMITS OF A CONNECTED SAMPLE

- Collected conversations measure activity, not everything a community does.
- Reading without posting leaves no trace in the data.
- A connected sample describes the network we followed, not the state we live in.

Collected conversations measure activity, not everything a community does.

- collected conversations versus everything a community does

- APIs expose activity, not attention

Reading without posting leaves no trace in the data.

- reading leaves no trace in the data

- the silent audience stays unmeasured

A connected sample describes the network we followed, not the state we live in.

- the snowball found its own network by design

- a bigger snowball is still a snowball
 - doubling the sample deepens the network, not the representativeness

The whole investigation

CLAIMS SIZED TO THE EVIDENCE THAT CARRIES THEM

- Close reading interprets particular posts and conversations. Distant reading finds patterns across whole corpora.
- The analysis climbed between them, and each move changed what could be claimed.
- Policy answers needed both: patterns to see the environment, close reading to say what the patterns meant.
- A claim's scale must fit the scale and construction of its evidence.

posts → conversations → accounts → networks → corpora → platform landscapes → the population (beyond this evidence)

Close reading interprets particular posts and conversations. Distant reading finds patterns across whole corpora.

- the bottom of the map is close reading, the top is distant

- the same evidence, two different kinds of attention

The analysis climbed between them, and each move changed what could be claimed.

- down for meaning, up for patterns

- one government question became two corpora, bounded findings and five areas for action, each stage a move on the map

Policy answers needed both: patterns to see the environment, close reading to say what the patterns meant.

- neither mode alone could answer the government’s three questions

- the report pairs patterns with interpretation before it recommends

A claim’s scale must fit the scale and construction of its evidence.

- the rule outlives this case

- fit claims to evidence wherever you argue

Displayed scale map

- the complete map, one last walk from posts to the population boundary

The shitposting milieu

WEEK 4'S SOCIAL STEGANOGRAPHY, AT COMMUNITY SCALE

- These communities speak in irony, in-jokes and provocation, a style often called shitposting.
- Week 4 called this social steganography: insiders read the signal, outsiders cannot settle the meaning.
- That is one reason posts alone cannot reveal intent.

These communities speak in irony, in-jokes and provocation, a style often called shitposting.

- provocation as vernacular, sincerity deniable

Week 4 called this social steganography: insiders read the signal, outsiders cannot settle the meaning.

- insiders decode, outsiders guess

That is one reason posts alone cannot reveal intent.

- deniability by design frustrates inference

- one more reason for careful, bounded claims

References

- Ballsun-Stanton, B., Waldek, L., Droogan, J., Smith, D., Iqbal, M., & Peucker, M. (2020). *Mapping Networks and Narratives of Online Right-Wing Extremists in New South Wales* (Version 1.0.1). Macquarie University. <https://doi.org/10.5281/zenodo.4071472>
- Moretti, F. (2000). Conjectures on world literature. *New Left Review*, 2(1), 54–68. <https://doi.org/10.64590/hxj>
- Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on 15 March 2019. (2020, December 8). 2. *The terrorist attack*. <https://christchurchattack.royalcommission.nz/the-report/part-1-purpose-and-process/the-terrorist-attack>

AI Use Disclosure

- Claude (Anthropic) and Codex (OpenAI) produced more than 1.2 million output tokens between 22 and 30 August.
- More than half of the output came from human-steered main threads.
- Two machines measured the work and their counts partly overlap.
- The tools supported source research, planning, drafting, production and verification.
- Brian selected the argument, sources and teaching form.

Speaker notes

Claude (Anthropic) and Codex (OpenAI) produced more than 1.2 million output tokens between 22 and 30 August.

- measured on two machines with the same estimator and window
 - office machine total of 1,143,742 output tokens, measured 31 August
 - home machine total of 647,689 output tokens, measured 30 August at 22:54
- the combined unique total is bounded, not exact
 - Codex sessions migrated between machines were counted on both
 - the bound runs from about 1.23 million to about 1.79 million
- the delivered total exceeds the measurement
 - builds after each measurement and the crash-recovery session add output

More than half of the output came from human-steered main threads.

- office measurement of 992,041 main and 151,701 subagent tokens

- home measurement of 280,561 main and 367,128 subagent tokens

- the main-thread share stays above half under every overlap value
 - delegated reads, reviews, edits and builds make up the rest

Two machines measured the work and their counts partly overlap.

- Week 6 development ran on the office machine from 22 August
 - a hard crash on 30 August moved the evening's work to the home machine

- migrated Codex sessions appear in both counts
 - the exact overlap is not recoverable from either machine alone
 - the totals are stated as bounds rather than a single figure

The tools supported source research, planning, drafting, production and verification.

- source and claim checking
 - live Zotero verification of quotations and page numbers
 - external Codex reviews of the goals layer, outline and sources

- teaching production
 - outline layering, slide composition and rendering
 - build validation and package checks

Brian selected the argument, sources and teaching form.

- Brian making the substantive decisions
 - every outline layer accepted before the next was drafted
 - content warning, support slide and tutorial rulings

- the lecturer retaining final responsibility
 - the report under discussion is Brian's own co-authored work
 - rendered artefacts reviewed before release

